CDMTCS
Research
Report
Series

Algebraic Constraints, Automata, and Regular Languages

Bakhadyr Khoussainov
Department of Computer Science
University of Auckland

CDMTCS-126
March 2000

Centre for Discrete Mathematics and
Theoretical Computer Science

Algebraic Constraints, Automata, and Regular Languages

Bakhadyr Khoussainov

Abstract

A class of decision problems is Boolean if it is closed under the set–theoretic operations of union, intersection and complementation. The paper introduces new Boolean classes of decision problems based on algebraic constraints imposed on transitions of finite automata. We discuss issues related to specifications of these classes from algebraic, computational and proof–theoretic points of view.

Key Words: Automata, Algebra, Computably Enumerable Sets, Σ_1–Algebra, Π_1–Algebra, Homomorphism, Congruence Relation, Isomorphism.

1 Introduction

Natural classes of decision problems usually possess closure properties under certain well-known operations. For example, the class of Turing decidable languages is closed under the operations of union, intersection, complementation, concatenation, and the star operation. So is the class of regular languages. Of course, not all known classes of decision problems are closed under the listed operations above. However, arguably Boolean classes, that is the classes closed under the operations of union, intersection and complementation, can be considered as natural classes of problems.

Suppose that we are given a class of decision problems. In the theory of formal languages, a traditional question that arises about the class is whether the class can be specified in an appropriate terminology. For example, the class of regular languages can be specified as the class of languages accepted by finite automata. Similarly, the class of all pushdown automata recognizable languages is specified as the context free grammars. There has also been research in characterzing other known classes of decidable problems, e.g. classes of problems decidable in polynomial time, using some formal systems of logic.

The primary goal of this paper is to introduce new Boolean classes of decision problems and discuss the issues related to the problem of specification of these classes. Each of these classes consists of regular langauges and is defined in terms of certain algebraic constraints on automata that recognize the languages.

Our motivation in introducing new Boolean classes of regular languages is twofold. The first motivation comes from a computational point of view. A run of a computer system can be thought as a sequence of states. During the run the system must satisfy certain constraints specified by the system software or/and hardware. Some of these constraints are of algebraic nature. For example, during the run, two consecutive executions of a program instruction I can produce the same result obtained by an execution of one program instruction J. This can algebraically be presented as an equation $II = J$. Constraints that force two program instructions, say I and J, to be executed in parallel, can also be presented as an algebraic equation $IJ = JI$. More generally, an algebraic expression of the type $I = J \to T = S$, can be understood as an algebraic constraint with the following meaning. Whenever the executions of program instructions I and J produce the same result then the results of executions of T and S are the same. These considerations suggest the idea of studying finite automata, or in general models of computations, whose transition tables satisfy certain algebraic constraints of the types above. The second motivation to introduce new Boolean classes of regular expressions comes from an algebraic point of view. It is well known that any finite deterministic automaton can be considered as a finite unary algebra. Similarly tree automata can be viewed as finite universal algebras [2] [4]. Hence, the concepts of algebra, e.g. finitely presented algebra, free algebra, equations, conditional equations, can be used in the study of regular languages and their properties.

We assume that the reader is familiar with the basics of finite automata, tree automata, regular languages [6], view of finite automata and tree automata as finite algebras [2], basics from the theory of universal algebras, e.g. finitely presented algebra, free algebra, congruence relations [5]. In addition, we use some notions from computability theory [11], e.g. c.e. set, simple set, immune set; and computable algebra [9], e.g. Σ_1–algebra, Π_1–algebra. Many of these notions will be defined as needed. A related paper discussing complexity issues is [3].

2 Automata with Algebraic Constraints

In this section, using terminology from universal algebra, we recall definitions of finite automata and regular languages, and introduce automata with algebraic constraints. So, fix a signature $\sigma =< f_1, \ldots, f_n, c_1, \ldots, c_m >$, where $c_1, \ldots, c_m$ are constant, and $f_1, \ldots, f_n$ are function symbols. An **al-**

gebra $\mathcal{A}$ of this signature is a system $< A, f_1, \dots, f_n, c_1, \dots, c_m >$, where each f_i is an operation on A and each c_j is a constant that interpret the appropriate symbols of the signature[1]. The algebra is **finite** if its **domain** A is a finite set. The **terms** of σ are defined by induction: each variable x and a constant c_j are terms; if $t_1, \dots, t_k$ are terms and f is a k–ary function symbol then $f(t_1, \dots, t_k)$ is a **term**. The set G of **ground terms** is the set of terms without variables. Each ground term defines a **finite labelled tree**: the leaves of the tree are labelled with the constants, other nodes are labelled with the function symbols, and any node labelled with symbol f of arity k has exactly k immediate successors.

Definition 1 *A* **language** *is a subset of the set G of ground terms.*

If one identifies the ground terms with trees then any language can be thought as a set of trees. A basic notion of this paper is the following.

Definition 2 *A* **finite automaton** *is a pair $M = (\mathcal{A}, F)$ consisting of finite algebra $\mathcal{A}$ and the set $F \subseteq A$. The elements of A are* **states***, F is the* **set of final states** *, and the constants $c_1, \dots, c_m \in A$ are the initial states of M. The* **associated with M algebra** *is then $\mathcal{A}$.*

If $m = 1$ and all functions are unary then M can be thought as a standard deterministic finite automaton over the alphabet $\{f_1, \dots, f_n\}$.

Let t be a ground term and $M = (\mathcal{A}, F)$ be an automaton. The automaton M evaluates the term t in a natural way: it is simply the value of the term t in the algebra $\mathcal{A}$. Procedurally this can be thought as follows. Think of t as a labelled tree. The leaves of t are values of the constants of the signature in the associated algebra $\mathcal{A}$. These are the initial states of M. If a node of the tree is labelled with f and the values of the immediate successors of the node are states $s_1, \dots s_k$ then label the node with the state $f(s_1, \dots, s_k)$. Thus the automaton works from the leaves to the root of t, and labels the nodes with states of M. The root is labelled with the state which is the value of t in $\mathcal{A}$.

Definition 3 *The automaton $M = (\mathcal{A}, F)$* **accepts** *the ground term t if the value of t in $\mathcal{A}$ is in F. Let $L(M)$ be the set of all ground terms accepted by M. The language $L(M)$ is* **regular**.

[1] We abuse notation and denote the function (constant) symbols and their interpretations with the same letters.

Any regular language is a decidable langauge. It is known that the class of all regular languages is a Boolean class. Below, using the concept of algebraic constraint, we provide other examples of Boolean classes consisting of regular languages.

An **algebraic constraint** is a conditional equation, that is the universal closure of the formula of the type $t_1 = q_1 \ \& \ \dots \ \& \ t_n = q_n \rightarrow t = q$, where t_i, q_i, t and q are terms of the signature. Let C be a set of algebraic constraints and let $M = (\mathcal{A}, F)$ be an automaton.

Definition 4 *The algebra $\mathcal{A}$ assocaited with M is a C–**algebra** if it satisfies all the formulas from C. The automaton $M = (\mathcal{A}, F)$ is a C–**automaton** if the associated algebra $\mathcal{A}$ is a C–algebra. The language accepted by a C–automaton is a C–**language**. Define R_C to be the set of all C–languages.*

Now our goal is to study the class R_C of all C–languages for a fixed set C of algebraic constraints. Note that there are no conditions on C, in particular C can be infinite. By the definition above any language from R_C is regular. Also note that R_C always contains G and $\emptyset$. The standard constructions of automata for recognizing the union, intersection, and the complements of regular languages (see for example [4]) produces the following result:

Theorem 1 *The class R_C of all C–languages is Boolean.*$\Box$

Definition 5 *An **equational constraint** is the universal closure of the formula of the type $t = q$, where t and q are terms. Let E be a set of equational constraints. An E–**automaton** is an automaton $M = (\mathcal{A}, F)$ such that $\mathcal{A}$ satisfies E. The algebra $\mathcal{A}$ is an E–**algebra**. A language L is an E–**language** if L is accepted by an E–automaton. Let R_E be the class of all E–languages.*

The above theorem holds for E–languages as well:

Theorem 2 *The class R_E of all E–languages is Boolean.*$\Box$

Algebraic constraints are generally not well–behaved as equational constraints. To demonstrate this we give the following notions and results about minimal automata. Let $\mathcal{M} = (\mathcal{A}, F)$ be an automaton. A **homomorphism** of M onto an automaton $M_1 = (\mathcal{A}_1, F_1)$ is a mapping h from $\mathcal{A}$ onto $\mathcal{A}_1$ such that h preserves the basic operations and for all states $s \in A$, $s \in F$ if and only if $h(s) \in F_1$. Note that in this case M and M_1 accept the same language. Equational constraints are always preserved under homomorphisms while algebraic constraints are not. We will use this in the following result. Recall that a **minimal automaton** for a regular language L is the automaton with the fewest states that accepts L.

Theorem 3 *Let L be an E–language. Then a minimal automaton for L is unique and is an E–automaton.*

Proof. It is a known fact that any regular language L has a minimal automaton accepting it. Moreover, the automaton is unique up to isomorphism. Additionally, any automaton that accepts L can be homomorphically mapped onto the minimal automaton [4]. So let M_1 be the minimal automaton for L. Since L is an E–language there exists an E–automaton M that accepts L. Since M_1 is minimal, the automaton M_1 is a homomorphic image of M. Thus, M_1 is an E–automaton since equational constraints are preserved under homomorphisms. The theorem is proved.

A natural relation defined by the the set C of algebraic constraints is the following. Ground terms t and q are C–**equivalent** if the equality $t = q$ can be proved (in the first order logic) from C. We denote C–equivalent terms t and q by $t \sim_C q$. Thus,

$$\sim_C = \{(p, q) \mid \ E \text{ proves } p = q\}.$$

The following lemma follows immediately.

Lemma 2.1 *The relation $\sim_C$ is a computably enumerable relation with an oracle for C. In particular, if C is a decidable set of algebraic constraints then $\sim_C$ is a c.e. relation.*□

For a set C of algebraic cosntraints any C–language possesses a natural completeness property with respect to the relation $\sim_C$. Formally, a language L is C–**complete** if for all $t, q \in G$ the condition $t \in L$ and $t \sim_C q$ implies that $q \in L$. Thus, any C–complete language is a union of some $\sim_C$–equivalence classes. These considerations now imply the following.

Corollary 2.1 *Any C–language is C–complete.* □

One can also study the global structure of R_E–classes, that is, we study the relationship between different classes R_E. Consider the set

$$K = \{R_E \mid E \text{ is a set of equational constraints}\}.$$

Thus, we have a partially ordered set $\mathcal{K} = (K, \subseteq)$. We state the next theorem about this partially ordered set and leave its proof as an exercise:

Theorem 4 *The partially ordered set $\mathcal{K}$ forms a complete lattice, where for all $R_{E_1}, R_{E_2} \in K$ the meet $R_{E_1} \wedge R_{E_2}$ coincides with $R_{E_1} \bigcap R_{E_2}$ and equals to $R_{E_1 \cup E_2}$, and the join $R_{E_1} \vee R_{E_2}$ is the minimal R_E that contains both R_{E_1} and R_{E_2}.* $\square$

3 Characterizations by Algebras

Let E be a set of equational constraints. The set G of all ground terms can naturally be transformed into an algebra: for any functional symbol f of arity k and ground terms $t_1, \ldots, t_k$, set the value of f on $(t_1, \ldots, t_k)$ be $f(t_1, \ldots, t_k)$. The algebra $\mathcal{F}$ thus obtained is the **absolutely free algebra with generators** $c_1, \ldots, c_m$. Recall that an equvalence relation η on $\mathcal{F}$ is a **congruence** if for all $a_1, \ldots, a_k, b_1, \ldots, b_k \in G$ and a basic k–ary operation f, the condition $(a_1, b_1), \ldots, (a_k, b_k) \in \eta$ implies that $(f(a_1, \ldots, a_k), f(b_1, \ldots, b_k)) \in \eta$. The equivalence relation $\sim_E$ induced by the equational constraints E (see Section 2) is a congruence relation of $\mathcal{F}$. Factorizing $\mathcal{F}$ by $\sim_E$, we obtain the **initial algebra** $\mathcal{F}_E$ defined by E. The algebra $\mathcal{F}_E$ possesses several natural properties: Any algebra that satisfies E and whose generators are $c_1, \ldots, c_m$ is a homomorphic image of $\mathcal{F}_E$, and this propery defines $\mathcal{F}_E$ uniquely up to an isomorphism.

Definition 6 *The algebra $\mathcal{F}_E$ is called* **initial for the class** R_E.

¿From the mentioned properties of $\mathcal{F}_E$, we obviously obtain the following.

Lemma 3.1 *For any E–automaton $M = (\mathcal{A}, F)$, the algebra $\mathcal{A}$ is a homomorphic image of $\mathcal{F}_E$. Moreover, if $\mathcal{F}_{E_1}$ is isomorphic to $\mathcal{F}_{E_2}$ then $R_{E_1} = R_{E_2}$.* $\square$

This lemma suggests the idea to say that the class R_E is characterized by the isomorphism type of the initial algebra $\mathcal{F}_E$. This idea does not work because there exist nonisomorphic $\mathcal{F}_{E_1}$ and $\mathcal{F}_{E_2}$ such that $R_{E_1} = R_{E_2}$. We refine the idea of characterizing the class R_E by the isomorphism types of algebras by introducing the following new notions. Let $FH(\mathcal{A})$ be the set of the isomorphism types of all finite homomorphic images $\mathcal{A}$.

Definition 7 *Two algebras $\mathcal{A}$ and $\mathcal{B}$ are* **relative** *if $FH(\mathcal{A}) = FH(\mathcal{B})$.*

Thus relative algebras can not be distingushed from each other by their finite homomorphic images. Now we prove the following theorem.

Theorem 5 *Two classes R_{E_1} and R_{E_2} coincide if and only if the initial algebras $\mathcal{F}_{E_1}$ and $\mathcal{F}_{E_2}$ are relative.*

Proof. Assume that the initial algebras $\mathcal{F}_{E_1}$ and $\mathcal{F}_{E_2}$ are relative. Take any language $L \in R_{E_1}$. There exists an E_1–automaton $M = (\mathcal{A}, F)$ that accepts the language. Then $\mathcal{A}$ is a homomorphic image of $\mathcal{F}_{E_1}$. Hence $\mathcal{A}$ must be a homomorphic image of $\mathcal{F}_{E_2}$ as well. We conclude that L is an E_2–language. Assume now that $R_{E_1} = R_{E_2}$. Suppose, without loss of generality, that there exists a finite homomorphic image $\mathcal{A}$ of $\mathcal{F}_{E_1}$ which does not belong to the set $FH(\mathcal{F}_{E_2})$. This implies that there exists an equation $t(x_1, \ldots, x_k) = q(x_1, \ldots, x_k)$ that is not satsified in $\mathcal{A}$ such that the equation belongs to E_2. Let $a_1, \ldots, a_k$ be elements in $\mathcal{A}$ that make this equation false in $\mathcal{A}$. There exist ground terms $p_1, \ldots, p_k$ such that a_i equals to the value of the term p_i in $\mathcal{A}$. Let $F = \{t(a_1, \ldots, a_n)\}$. Consider the automaton $(\mathcal{A}, F)$. This automaton accepts the ground term $t(p_1, \ldots, p_k)$, but does not accept the term $q(p_1, \ldots, p_k)$. Since $R_{E_1} = R_{E_2}$ it must be the case that $L \in R_{E_2}$ since L is accepted by an E_1–automaton. But then L is E_2–complete by Corollary 2.1. Therefore $q(p_1, \ldots, p_k)$ must belong to L since the equality $t(x_1, \ldots, x_k) = q(x_1, \ldots, x_k)$ belongs to E_2. This is a contradiction. The theorem is proved. □

Definition 8 *An algebra $\mathcal{A}$ is a* **character** *of the class R_E if $FH(\mathcal{A})$ consists of all algebras associated with E–automata.*

Thus, for example the algebra $\mathcal{F}_E$ and, by the theorem above, any algebra $\mathcal{F}_{E'}$ that is relative to $\mathcal{F}_E$ are characters of the class R_E.

Lemma 3.2 *Any algebra is a character for some class R_E.*

Proof. Let $\mathcal{A}$ be an algebra. Consider the set $E(\mathcal{A})$ of all equations satisfied by $\mathcal{A}$. Then the algebra $\mathcal{A}$ is the initial algebra defined by $E(\mathcal{A})$. Therefore the algebra $\mathcal{A}$ is a character of the class $R_{E(\mathcal{A})}$. This proves the lemma. □

Corollary 3.1 *Any two relative algebras are characters of the same class of regular languages. Particularly for any E, the initial algebra $\mathcal{F}_E$ and any algebra relative to $\mathcal{F}_E$ are characters of the class R_E.* □

For a given set E of equational constraints, consider the set $Ch(R_E)$ of all isomorphism types of algebras relative to $\mathcal{F}_E$. A natural question is whether one can define an algebra from this set which, in certain sense, is a canonical character for R_E. One way to do this is the following. On $Ch(R_E)$ introduce the relation $\leq_h$: for all $\mathcal{A}, \mathcal{B} \in Ch(R_E)$, $\mathcal{A} \leq_h \mathcal{B}$ iff there exists a homomorphism from $\mathcal{B}$ onto $\mathcal{A}$. This relation is a partial order. The next theorem shows that $(Ch(R_E), \leq_h)$ has a unique minimal element. So one can say that the minimal element is a canonical character of the class R_E.

Theorem 6 *For any R_E there exists a character $\mathcal{C}_E$ of the class R_E such that every character of the class R_E is homomorphically mapped onto $\mathcal{C}_E$.*

Proof. Consider the absolutely free algebra $\mathcal{F}$. Consider the class of all algebras associated with E-automata. This class coincides with the class of all finite homomorphic images of $\mathcal{F}_E$. Let $\mathcal{A}_0, \mathcal{A}_1, \ldots$ be a list of all such finite algebras. Define the following equivalence relation $\sim_E^r$ on the set GT of ground terms: two terms t and q are $\sim_E^r$–equivalent, written $t \sim_E^r q$, if in the algebra $\mathcal{A}_i$ the equality $t = q$ holds for all i. One now checks that $\sim_E^r$ is a congruence relation on $\mathcal{F}$. Hence factorizing $\mathcal{F}$ by $\sim_E^r$, we obtain the algebra which we denote by $\mathcal{C}_E$. We want to show that $\mathcal{C}_E$ satsifies the properties stated by the theorem. First we show that $\mathcal{C}_E$ is relative to the initial algebra $\mathcal{F}_E$. Let $\mathcal{B}$ be a finite algebra from $FH(\mathcal{F}_E)$. We define a mapping h from $\mathcal{C}_E$ to $\mathcal{B}$ as follows. Take an $a \in C_E$. There exists a ground term t whose value in $\mathcal{C}_E$ equals to a. Let b be the value of the ground term t in the algebra $\mathcal{B}$. Then, one can check that the mapping $h(a) = b$ is a homomorphism from $\mathcal{C}_E$ onto $\mathcal{B}$. Now we want to show that any finite homomorphic image of $\mathcal{C}_E$ is also a homomorphic image of $\mathcal{F}_E$. It suffices to show that $\mathcal{C}_E$ is a homomorphic image of $\mathcal{F}_E$. Since $\mathcal{F}_E$ is the initial algebra for E, it suffices to prove that any equality $t = q$ between ground terms that is true in $\mathcal{F}_E$ is also true in $\mathcal{C}_E$. Let $t = q$ be an equality between ground terms that are true in $\mathcal{F}_E$. Then $t = q$ holds in every finite algebra $\mathcal{A}_i$. Hence, by the definition of $\sim_E^r$, the terms t and q are $\sim_E^r$–equivalent. Hence $t = q$ is true in $\mathcal{C}_E$. Therefore $\mathcal{C}_E$ is, in fact, a homomorphic image of $\mathcal{F}_E$. Hence any finite homomorphic image of $\mathcal{C}_E$ is also a homomorphic image of $\mathcal{F}_E$. This shows that $\mathcal{C}_E$ and $\mathcal{F}_E$ are relative algebras.

In order to prove the second part of the theorem we need to show that any algebra $\mathcal{B}$ relative to $\mathcal{F}_E$ can be homomorphically mapped onto $\mathcal{C}_E$. Let b be an element of $\mathcal{B}$. Take a term t whose value in $\mathcal{B}$ is b. Map b onto the value of the term t in $\mathcal{C}_E$. This mapping does not depend on the choice of

t. Hence there exists a homomorphism from $\mathcal{B}$ onto $\mathcal{C}_E$. The theorem is proved. □

Definition 9 *The* **canonical character** *of the class R_E of decision problems is the algebra $\mathcal{C}_E$ which is the minimal element of $(Ch(R_E), \leq_h)$.*

The next section studies some computational properties of the canonical characters for certain classes of R_E. The section provides a necessary and sufficient condition for the canonical character of R_E to coincide with the initial algebra $\mathcal{F}_E$.

4 On Canonical Characters of R_E

All the characters of the class R_E of decision problems that satisfy E are among homomorphic images of the algebra $\mathcal{F}_E$. Thus, the partially ordered set $(\{A|\mathcal{A} \leq_h \mathcal{F}_E\}, \leq_h)$ has the minimal element $\mathcal{C}_E$ and the maximal element $\mathcal{F}_E$. In this section we find conditions when $\mathcal{F}_E$ coincides with $\mathcal{C}_E$, and study some computability–theoretic properties of the canonical characters. To do this, we need to introduce a couple of notions from universal and computable algebra.

Definition 10 *An algebra $\mathcal{A}$ is* **residually finite** *if for all $a, b \in A, a \neq b$ there is a homomorphism h of $\mathcal{A}$ onto a finite algebra such that $h(a) \neq h(b)$.*

The notion of residually finite algebra is a fundamental concepts of universal algebra and plays an important role classifying certain algebraic structures[5]. Now we introduce standard notions from computable algebra. Consider an algebra $\mathcal{A}$ of the signature σ. There is a congruence relation η on $\mathcal{F}$ such that $\mathcal{A}$ is isomorphic to the algebra obtained by factorizing $\mathcal{F}$ by η.

Definition 11 *The algebra $\mathcal{A}$ is a* Π_n**–algebra** *(*Σ_n**–algebra***) if the relation η is a Π_n–set (Σ_n–set). If $\mathcal{A}$ is both a Σ_1–algebra and Π_1–algebra then $\mathcal{A}$ is a* **computable algebra**.

Examples of Σ_1–algebras are the initial algebras $\mathcal{F}_E$ for computably enumerable sets of constraints E. In the theory of computable algebras there has been an extensive interest in Σ_1–algebras while there has not been an emphasis in the study of Π_1–algebras because of the small number of natural examples. It turns out that canonical characters are the source of natural examples of Π_1–algebras. Here is a simple result.

Lemma 4.1 *If the class of all finite homomorphic images of $\mathcal{F}_E$ is computably enumerable then the canonical character $\mathcal{C}_E$ for the class R_E is a* Π_1*–algebra.*

Proof. By the assumption, there exists a sequence $\mathcal{A}_0, \mathcal{A}_1, \ldots$ of all finite homomorphic images of $\mathcal{F}_E$ such that the set $\{(x, y)|x \in A_i\}$ is c.e.. Consider the the congruence $\sim^r_E$ that defines the canonical algebra $\mathcal{C}_E$. By the definition, $t \sim^r_E q$ iff for $\forall i(t = q \text{ in } \mathcal{A}_i)$. Hence, η is a Π_1–relation. Therefore, the algebra $\mathcal{C}_E$ is a Π_1–algebra. The lemma is proved. □

Corollary 4.1 *For any finite set E, the canonical character $\mathcal{C}_E$ for the class R_E is a* Π_1*–algebra.*

Proof. The set E is finite. So, effectively list all finite algebras that satisfy E. These algebras are homomorphic images of $\mathcal{F}_E$. Hence the hypothesis of the lemma above holds true.□

The next theorem gives a criterium as when the partially ordered set $(\{\mathcal{A}|\mathcal{A} \leq_h \mathcal{F}_E\}, \leq_h)$ has a unique element, that is when $\mathcal{F}_E = \mathcal{C}_E$.

Theorem 7 *For a given class R_E of decision problems, the initial algebra $\mathcal{F}_E$ is residually finite if and only if the algebras $\mathcal{F}_E$ and $\mathcal{C}_E$ coincide.*

Proof. Consider the class R_E. Assume that $\mathcal{F}_E$ is a residually finite algebra. We want to show that the minimal character $\mathcal{C}_E$ for the class R_E is isomorphic to $\mathcal{F}_E$. From the proof of Theorem 6, we know that $\mathcal{C}_E$ is a homomorphic image of $\mathcal{F}_E$. Let h be the homomorphism. We want to show that h is a one to one mapping. Indeed, let a, b be two distinct elements in F_E. Then, there exist ground terms $t(p_1, \ldots, p_k)$ and $q(r_1, \ldots, q_s)$ such that the values of these terms in the algebra $\mathcal{F}_E$ are a and b, respectively. Since $\mathcal{F}_E$ is a residually finite algebra there exists a finite homomorphic image $\mathcal{A}_i$ of $\mathcal{F}_E$ in which the images of a and b are also distinct. Therefore the ground terms $t(p_1, \ldots, p_k)$ and $q(r_1, \ldots, r_s)$ are not $\sim^r_E$–equivalent, where $\sim^r_E$ is the congruence relation that defines the algebra $\mathcal{C}_E$. Hence the mapping h must be a one to one mapping since $h(a) \neq h(b)$ by the definition of $\sim^r_E$.

Assume now that $\mathcal{F}_E$ and the minimal character $\mathcal{C}_E$ coincide. For the sake of construdiction, also assume that $\mathcal{F}_E$ is not residually finite. Hence there exist two distinct elements a and b in $\mathcal{F}_E$ such that in any finite homomorphic image of $\mathcal{F}_E$ the images of a and b are equal. Let $t(p_1, \ldots, p_k)$ and $q(r_1, \ldots, r_s)$ be ground terms whose values in $\mathcal{F}_E$ are a and b, respectively.

Then the images of these elements in any finite homomorphic image of $\mathcal{F}_E$ are equal. Therefore, by the definition of the equivalence relation $\sim_E^r$, the ground terms $t(p_1, \ldots, p_k)$ and $q(r_1, \ldots, r_s)$ must be equal in the algebra $\mathcal{A}$. But this is not possible because $\mathcal{A}$ and $\mathcal{F}_E$ coincide. Contradication. The theorem is proved. □

Corollary 4.2 *For any finite set E, if the initial algebra $\mathcal{F}_E$ is residually finite then the minimal character $\mathcal{C}_E$ of the class R_E is a computable algebra.*

Proof. The initial algebra $\mathcal{F}_E$ and the canonical character $\mathcal{C}_E$ are isomorphic. Since E is finite, $\mathcal{F}_E$ is a Σ_1–algebra. By Corollary 4.1 the canonical character $\mathcal{C}_E$ is a Π_1–algebra. So, the algebra $\mathcal{F}_E$ is both a Σ_1–algebra and Π_1–algebra. Hence $\mathcal{F}_E = \mathcal{C}_E$, and $\mathcal{C}_E$ is a computable algebra. □

5 Finite Equational Constraints

In the previous two sections we introduce the notion of character as a tool to specify a given class R_E. This is an algebraic approach to the specification problem of the class R_E. One can study the specification problem from computational and logical point of view as well. From computational point of view it is quite natural to be interested in finiding a finite E' such that $R_E = R_{E'}$. However, we have already noted there could be an $E' \neq E$ such that E and E' have different proof–theoretic power with $R_E = R_{E'}$. Therefore from logical (proof–theoretic) point of view point it is natural to be interested in finding a finite E' for which $\sim_E = \sim_{E'}$. Of course, in this case we have $R_E = R_{E'}$. These observations lead us to the following definitions.

Definition 12 *The pair (R_E, E) has a* **a finite specification** *if there exists a finite E' for which $\sim_E = \sim_{E'}$. The class R_E has a* **finite specification** *E' if $R_E = R_{E'}$ and E' is finite.*

If (R_E, E) has a finite specification then clearly R_E has a finite specification. Below is a theorem that gives examples of classes that have finite specifications. But we first prove the following lemma:

Lemma 5.1 *Let $\mathcal{A}$ be a finite algebra, and $E(\mathcal{A})$ be the set of all equations satisfied by $\mathcal{A}$. Then the pair $(R_{E(\mathcal{A})}, E(\mathcal{A}))$ has a finite specification.*

Proof. To prove the lemma we introduce the notion of hight $h(t)$ for ground terms t. The hight of any constant c is 0. If the hights

$h(t_1), \ldots, h(t_m)$ have been defined, then $h(f(t_1, \ldots, t_m)) = max\{h(t_i) \mid i = 1, \ldots, m\} + 1$. Since the algebra $\mathcal{A}$ is finite there exists a minimal s such that every term of hight s equals, in the algebra $\mathcal{A}$, to a term whose hight is less than s. The number of terms of hight $\leq s$ is finite. Let $E' = \{t = q | h(t), h(q) \leq s$ the algebra $\mathcal{A}$ satisfies $t = q\}$. Note that E' is finite. Now $\mathcal{F}_{E'}$ is isomorphic to the algebra $\mathcal{A}$. Therefore $\sim_{E(\mathcal{A})} = \sim_{E'}$. The lemma is proved.

Theorem 8 *For any finite set X of regular languages, the minimal class $R(X) \in K$ that contains X has a finite specification.*

Proof. Let $X = \{L_1, \ldots, L_k\}$. Consider the minimal automaton $M_i = (\mathcal{A}_i, F_i)$ that accepts L_i, $i = 1, \ldots, k$. Consider the congruence relation η_X on $\mathcal{F}$ defined as follows: $(t, q) \in \eta_X$ iff $t = q$ in $\mathcal{A}_i$ for $i = 1, \ldots, k$. Let $\mathcal{F}(X)$ be the algebra obtained by factorizing $\mathcal{F}$ by η_X. The algebra $\mathcal{F}(X)$ is the minimal algebra with respect to $\leq_h$ in the class of all algebras $\mathcal{A}$ such that $\{\mathcal{A}_1, \ldots, \mathcal{A}_k\} \subset FH(\mathcal{A})$. Therefore $R(X)$ coincides with the class of all regular langauges that are accepted by automata whose associated algebras belong to $FH(\mathcal{F}(X))$. Note that $\mathcal{F}(X)$ is isomorphic to the Cartesian product $\mathcal{A}_1 \times \ldots \times \mathcal{A}_k$ because $(t, q) \in \eta_X$ if and only if $t = q$ holds in $\mathcal{A}_1 \times \ldots \times \mathcal{A}_k$. Thus, from the lemma above we conclude that the theorem is proved. □

Lemma 5.2 *If the pair (R_E, E) has a finite specification E' then the algebra $\mathcal{F}_E$ is a Σ_1–algebra. Moreover, if $\mathcal{F}_E$ is residually finite then $\mathcal{F}_E$ is a computable algebra.*

Proof. The algebras $\mathcal{F}_E$ and $\mathcal{F}_{E'}$ are isomorphic. Since E' is finite, the congruence relation $\sim_{E'}$ is a c.e. relation. Hence the algebra $\mathcal{F}_E$ is a Σ_1–algebra. If $\mathcal{F}_E$ is residually finite then, by Corollary 4.2, the algebra $\mathcal{F}_E$ is computable. □

Corollary 5.1 *If (R_E, E) has a finite specification and $\mathcal{F}_E$ is not computable then $\mathcal{F}_E$ is not residually finite.* □

The results above lead us to the following question. If the initial algebra $\mathcal{F}_E$ is a computable and residually finite, does then the pair (R_E, E) has a finite specification? The theorem below answers the question.

Theorem 9 *There exists an E such that $\mathcal{F}_E$ is computable and residually finite but the pair (R_E, E) does not have a finite specification.*

Proof. Consider the signature is $< f_1, f_2, c >$, where f_1, f_2 are unary function symbols. Define the congreuce relation η on $\mathcal{F}$ as follows: $t\eta q$ iff $t = q$ or $h(t) = h(q) = 2^n$ for some n. The algebra $\mathcal{A}$, obtained by factorizing $\mathcal{F}$ by η, is computable. Moreover, $\mathcal{A}$ is residually finite. Consider $E = E(\mathcal{A})$, the set of all equations true in $\mathcal{A}$. We claim that the pair (R_E, E) does not have a finite specification. To show this we analyse the equations true in $\mathcal{A}$. Let the universal closure of the equation $t = q$ be true in $\mathcal{A}$. Then $h(t) = h(q)$. Suppose t and q contain variables x and y, respectively. So we write $t(x)$ and $q(y)$ instead of t and q. Then $x = y$, otherwise the equation would not be true in $\mathcal{A}$. We claim, if $x = y$ then $t = q$. Otherwise, let n be such that $m = 2^n > h(t)$. Then, since the universal closure of $t(x) = q(x)$ is true in $\mathcal{A}$, the $t(f_1^m(c)) = q(f_1^m(c))$ is also true in $\mathcal{A}$. By the definition of $\mathcal{A}$, this is not possible. Also, it is not the case that only one of the terms t, q contains a variable. Now assume that for some finite E' we have $\mathcal{F}_E = \mathcal{F}_{E'}$. Then, we can assume that no equation $t = q$ in E' contains a variable. Set $s = max\{h(t)|t = q \in E' \text{ or } q = t \in E'\}$. Let $r = 2^s$. Then the equality $f_1^r(c) = f_2^r(c)$ can not be derived from E'. This is a contradiction. The theorem is proved. □

Corollary 5.2 *There exists a class R_E such that the canonical character $\mathcal{C}_E$ is computable, residually finite but R_E does not have a finite specification.*

Proof. Consider E defined in the theorem above. Assume that for some finite E', $R_E = R_{E'}$. Then it must be the case that $\sim_{E'} \subset \sim_E$. Hence for all $(t, q) \in E'$, the hight of t equals to the hight of q. Since E' is finite there exist two terms t, q such that $(t, q) \notin E'$ but $(t, q) \in E$. Then there exists a homomorphic finite image of $\mathcal{F}_{E'}$ in which t and q are also distinct. Hence $\mathcal{F}_{E'}$ is not relative to $\mathcal{F}_E$. Therefore $R_E \neq R_{E'}$ by Theorem 5. The corollary is proved. □

Theorem 9 and the corollary above show that it is not always possible to find a finite specification for a class R_E even when the initial algebra $\mathcal{F}_E$ is computable and residually finite. This suggests the idea to refine the notion of finite specification. We do this by considering expansions of the original signature. An **expansion** of the signature σ is obtained by adding finitely many new function symbols. If $\mathcal{A}$ is an algebra of σ then by taking interpretations of the new function symbols in A, we obtain an **expansion** of $\mathcal{A}$. Then $\mathcal{A}$ is called a σ**–reduct** of the expansion.

Let σ_1 be an expansion of the signature σ. Let E, E_1 be sets of algebraic constraints of the signatures σ, σ_1 respectively. We say R_{E_1} is a **refinement** of R_E if R_{E_1} is infinite and the σ–reduct of any E_1–algebra is an E–algebra.

Definition 13 *We say that R_E is* **expansionary specified** *if there exists a refinement R_{E_1} such that R_{E_1} has a finite specification.*

¿From this definition the following proposition follows easily.

Proposition 1 *If R_E is expansionary specified then there exists a Σ_1–algebra which is initial for some refinement of R_E.* $\square$

In the next section we show that the converse of this proposition does not hold true.

6 A Counterexample

In this section we fix the signature $\sigma =< f_1, f_2, c >$, where f_1, f_2 are unary function symbols. We need some notions from computability theory. An infinite subset of the set G of ground terms is **immune** if it contains no infinite c.e. subsets. A c.e. set $X \subset G$ with immune complement $\bar{X}$ is called **simple**. A set $X \subseteq G$ is a **weak subalgebra** if $f_1(x), f_2(x) \in X$ for all $x \in X$. If $c \in X$ for a weak subalgebra X then $X = G$.

Lemma 6.1 *There exists a simple weak subalgebra of $\mathcal{F}$.*

Proof. Let $W_0, W_1, \dots$ be a standard enumeration of all c.e. subsets of G. We construct X by stages. At stage s we define a set X_s, then put $X = \bigcup_S X_s$. Constructing X, we need to satisfy the requirement r_i stating that $W_i \bigcap \bar{X} \neq \emptyset$ for all i such that W_i is infinite and $W_i \not\subset X$. For any $Y \subset G$, let $Cl(Y)$ be the set of all terms containing subterms of Y. Say that r_i **attracts attention at** s if $W_{i,s} \bigcap X_s = \emptyset$ and $W_{i,s} \neq \emptyset$. We set $X_0 = \emptyset$.

Stage s. Assume that X_{s-1} has been constructed. Find the minimal r_i, $i \leq s$, that requires attention. Take the first term $t \in W_{i,s}$ such that $h(t) > i + 1$, and set $X_{s+1} = Cl(X_s \bigcup \{t\})$. Go to the next stage. If no $i \leq s$ requires attention then go to the next stage.

The set X is c.e.. Note that for each i there is a term $t \notin X$ of length $i + 1$. Hence $\bar{X}$ is infinite. If X is not simple, then take the minimal i for which $W_i \subset \bar{X}$ and W_i is infinite. Consider the stage t, after which no r_j, $j < i$, requires attention. Then there is a stage $s > t$ at which r_i requires attention. Hence $W_{i,s} \bigcap X_s \neq \emptyset$, and therefore $W_i \bigcap X \neq \emptyset$. So X is simple. By the construction X is a weak subalgebra. The lemma is proved.$\Box$

Theorem 10 *There exists a class R_E with no expansionary specifications such that $\sim_E$ is a computably enumerable set.*

Proof. Consider the free algebra $\mathcal{F}$ of the signature $\sigma =< f_1, f_2, c >$. Consider the set X constructed in the lemma above. Define a c.e. congruence relation η: $(t, q) \in \eta$ iff $t = q$ or $t, q \in X$. Now let $\mathcal{A}$ be the Σ_1–algebra obtained by factorizing $\mathcal{F}$ by η. Set $E = E(\mathcal{A})$, where $E(\mathcal{A})$ is the set of all equations true in $\mathcal{A}$. We show that R_E is the required class.

Let f be a basic n–ary operation of an algebra $\mathcal{B}$. A **transition of** $\mathcal{B}$ is any of the mappings $f(a_1, \ldots, a_{n-1}, x)$, $\ldots$, $f(x, a_1, \ldots, a_{n-1})$, where $a_1, \ldots, a_{n-1} \in B$ are fixed. Let $Tr(\mathcal{B})$ be the algebra whose basic operations are all transitions of $\mathcal{B}$. Then a relation α is a congruence relation of $\mathcal{B}$ if and only if α is a congruence of $T(\mathcal{B})$.

Let $\mathcal{A}' = (A', f_1, \ldots, f_n)$ be a Σ_1-expansion of the initial algebra of some refinmenet of R_E.

Let X' be the image of X in $\mathcal{A}'$. Note that X' is infinite and is a simple set. We want to show that $\mathcal{A}'$ is is residually finite. Let $F_0, F_1, \ldots$ be an effective list of the transitions all the transitions of the expanded algebra $\mathcal{A}'$. Consider **the transition algebra of** $Tr(\mathcal{A}')$. As noted above, it suffices to prove that $Tr(\mathcal{A}')$ is residually finite.

Let $t_1, t_2 \notin X'$ be two distinct ground terms in $\mathcal{A}'$. We will show that there exists a finite set $S \subset \bar{X}'$ such that $t_1, t_2 \in S$ and the relation $eq(S) = \{(x, y) | x, y \in G \backslash S\} \bigcup \{(x, y) | x = y\}$ induces a congruence relation of $Tr(\mathcal{A}')$.

If a such S exists, then the mapping $h : t \rightarrow \{s | (t, s) \in eq(S)\}$ will be a homomorphism from $\mathcal{A}'$ onto a finite algebra in which $h(t_1) \neq h(t_2)$. In order to prove that there exists a set S with the above three properties we need to make several notes. Take a term $x \in X'$, a transition F_i, and a finite $S' \subset \bar{X}'$. If $F_i(x) \notin S'$ then $\{t | F_i(t) \in S'\} \subset \bar{X}'$. This set is computable and hence since $\bar{X}$ is immune. If $F_i(u) \in S'$ then $F_i(q) = F_i(x)$ for all $q \in X'$, and again $\{t | F_i(t) \neq F_i(u) \text{in } \mathcal{A}'\} \subset \bar{X}'$. This set is computable and hence finite. Also note the following fact. The equivalence relation $eq(S) = \{(x, y) | x, y \in G \setminus S\} \bigcup \{(x, y) | x = y\}$ is a congruence for the transition F_i if and only if for all ground terms $t' \notin S$, we have $F_i(x) \in S \longleftrightarrow F_i(t') = F_i(x)$, and $F_i(x) \notin S \longleftrightarrow F_i(t) \notin S$. Now we give a stagewise construction of S. At stage 0 we put $S_0 = \{t_1, t_2\}$. Clearly $S_0 \subset \bar{X}'$.

Stage j+1. Suppose that S_j has been constructed and $S_j \subset \bar{X}$. Consider the transitions $F_0, \ldots, F_{j+1}$. For each $i \leq j + 1$, consider $F_i(x)$.

If $F_i(x) \not\in S_j$, then set $S_{j+1,i} = S_j \bigcup \{t | F_i(t) \in S_j\}$. Otherwise, set $S_{j+1,i} = S_j \bigcup \{t | F_i(t) \neq F_i(x) \text{ in } \mathcal{A}'\}$. Define $S_{j+1} = S_{j+1,0} \bigcup \ldots \bigcup S_{j+1,j+1}$.

By the remarks given before the construction, the set $S = \bigcup_j S_j$ is a finite subset of $\bar{X}$. There exists a stage j_0 such that $S = S_{j_0}$. The terms t_1 and t_2 belong to S. We have to show that $eq(S)$ induces a congruence relation for every transition F_i. It suffices to prove that if s does not belong to S, then $(F_i(x), F_i(s)) \in eq(S)$. Consider any stage $j \geq j_0$. Suppose that $F_i(x) \not\in S_j$. Then $F_i(s) \not\in S_j$, otherwise $s \in S_j$ and hence $S_{j_0} \neq S_j$. Similarly, if $F_j(x) \in S_j$, then $F_j(s) = F_j(x)$, otherwise $s \in S_j$ and hence $S_{j_0} \neq S_j$. Thus, the homomorphism h defined by $h : t \rightarrow \{s | (t,s) \in eq(S)\}$ maps $\mathcal{A}$ onto a finite algebra in which $h(t_1) \neq h(t_2)$. Thus, $\mathcal{A}'$ is residually finite.

To finish the proof, assume that R_E has an expansionary specification. Then there exists a refinement of R_E that has a finite specification. Let $R_{E'}$ be a such refinement so that E' is finite. Let $\mathcal{A}'$ be the initial algebra $\mathcal{F}_{E'}$. Then $\mathcal{A}'$ is residually finite and hence is computable by Corollary 4.2. Therefore the set X is computable. This is a contradiction.

The theorem is proved.

References

[1] J.A. Bergstra, J.V.Tucker, Algebraic Specifications of Computable and Semicomputable Data Types, Theoretical Comp. Science, 50, 1987.

[2] J. R. Büchi. *Finite Automata, Their Algebras and Grammars: Towards a Theory of Formal Expressions*, D. Siefkes (editor), Springer-Verlag, 1989.

[3] M. Dinneen and B. Khoussainov. *Automata with Equational Constraints*, Submitted.

[4] F. Gécseg and M. Steinby. *Tree Automata*, Akadémiai Kiadó, Budapest, 1984.

[5] G. Grätzer. *Universal Algebra*, Springer-Verlag, New York and Heidelberg, 2nd edn., 1979.

[6] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation*, Addison–Wesley, 1979.

[7] A.I. Malcev. *Constructive Algebras*, Uspekhi Matem. Nauk, 16, No 3, 1961, 3-60.

[8] N.K. Kassimov, On Finitely Approximable and R.E. Representable Algebras, Algebra and Logic, 26, No 6,1986.

[9] *Handbook of Recurive Mathematics*, Volume 1. Studies in Logic and Foundations of Mathematics, edited by Yu. Ershov, S. Goncharov, A. Nerode, J. Remmel. Associate Editor V. Marek, Elsevier, 1998.

[10] B.Khoussainov. *Radomness, Computability, and Algebraic Specifications*, Annals of Pure and Applied Logic, 91, 1998, 1-15.

[11] R. Soare. *Recursively Enumerable Sets and Degrees.* The Study of Computable Functions and Computably Generated Sets. Perspect. in Math. Logic, 1987.